

Why should prediction discrepancies be renamed Standardized VPC ?

Authors: Emmanuelle Comets¹, PhD, Karl Brendel², PhD, and France Mentré³, MD, PhD

Address: ¹INSERM, UMR 738, F-75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, UMR 738, F-75018 Paris, France ; ² IRIS, Servier, Neuilly sur Seine, France; ³ Univ Paris Diderot, Sorbonne Paris Cité, UMR 738, F-75018 Paris, France; INSERM, UMR 738, F-75018 Paris, France

Keywords: non-linear mixed effect models, model evaluation, visual predictive check, prediction discrepancies, normalised prediction distribution errors

We have read the paper by Wang and Zhang¹. We are pleased to see that other authors further illustrate and also find nice properties for an evaluation tool that we first used more than ten years ago², presented at PAGE in 2000³ and published extensively first in 2006⁴. However we disagree with some of the conclusions made by the authors, and we feel that an important reference should have been included in their manuscript.

As Wang and Zhang point out themselves, what they call SVPC is nothing other than the prediction discrepancies (pd) named that way, after being called pseudo-residuals by Mentré and Escolano in 2006⁴. It is therefore misleading to present SVPC as something novel when in fact it goes back on something that our group has published and presented in conferences.

Plots of SVPC versus time in the paper by Wang and Zhang¹ are very similar to plots of pd versus time in Mentré and Escolano⁴, therefore we disagree with the statement on page 2 that "Neither pd nor npde was intended/recommended for evaluation of model predictions over a time course". pd and npde were developed for their improved statistical properties over linearisation-based residuals but are used as visual diagnostic tools in a similar way.

Also the paper is not complete when it comes to the current state of literature. Wang and Zang¹ missed one important reference, in which we have compared pd, npde, VPC, as well as tests based on prediction intervals with or without decorrelation⁵. We have also proposed tests for covariate models illustrating how npde can be used to evaluate covariate models. Their paper is rather inaccurate when comparing the properties of the pd to those of the npde. The idea behind the npde, as recalled in the paper, is to decorrelate observations within individuals. In Mentré and Escolano⁴, it had been shown that the type I error of the test based on pd was inflated when subjects contribute several observations, because of the within-subject correlation between observations, and it was anticipated that the decorrelation would improve this feature. Indeed, in Brendel et al.⁵ we have shown that the type I error of the npde is close to the expected 5% while the type I error for the pd or the test based on prediction intervals of the VPC was larger. In that paper, we performed a full simulation study, i.e. with several replications of the simulated data set, while in the manuscript of Wang and Zhang only one simulated dataset is given, from which it is very hard to draw meaningful conclusions. It is quite possible that for one simulated dataset the pd detects something that the npde does not, indeed, our simulations have shown that on a large number of simulated datasets, the tests with npde maintain type I error while the tests with pd have an increased type I error and hence a higher power. A discussion of the power to detect model misspecification should therefore take into account the increase in type I error under the null hypothesis. Also in their table V, there was a significant departure from 0 of the mean npde for the wrong model ($p=0.03$ with a Wilcoxon test).

We would also like to take this opportunity to mention that more informative VPC graphs have been proposed by Wilkins⁶ *et al* in 2006, which use prediction intervals around the simulated percentiles, and that we have adapted these graphs to pd and npde⁷. In that paper, we made a case

of using pd instead of npde to plot diagnostic graphs because the decorrelation tends to blur the relationship with time when used for visual diagnostics.

Finally, we join Wang and Zhang to stress the nice properties of simulation-based model evaluation tools such as pd (SVPC) over VPC and to encourage readers to use them.

References:

1. Wang D, Zhang S. Standardized Visual Predictive Check versus Visual Predictive Check for model evaluation. *J Clin Pharmacol*. 2011; *in press*.
2. Mesnil F, Mentré F, Dubruc C, Thénot JP, Mallet A. Population pharmacokinetics analysis of mizolastine and validation from sparse data on patients using the nonparametric maximum likelihood method. *J Pharmacokinet Biopharm*. 1998; 26: 133-61.
3. Mentré F, Escolano S. Validation methods in population pharmacokinetics: a new approach based on predictive distributions with an evaluation by simulation. *Annual Meeting of the Population Approach Group in Europe*, PAGE 9, 2000; Abstr 85.

<http://www.page-meeting.org/?abstract=85>

4. Mentré F, Escolano S. Prediction discrepancies for the evaluation of nonlinear mixed-effects models. *J Pharmacokinet Pharmacodyn*. 2006; 33:345-67.
5. Brendel K, Comets E, Laffont C, Mentré F. Evaluation of different tests based on observations for external model evaluation of population analyses. *J Pharmacokinet Pharmacodyn*. 2010 ; 37: 49-65.
6. Wilkins J, Karlsson M, Jonsson EN. Patterns and power for the visual predictive check. *Annual Meeting of the Population Approach Group in Europe*, PAGE 15, 2006; Abstr 1029.

<http://www.page-meeting.org/?abstract=1029>

7. Comets E, Brendel K, Mentré F. Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics. *Journal de la Société Française de Statistique*. 2010; 151: 1-24.

http://smf4.emath.fr/Publications/JSFDS/151_1/pdf/sfds_jsfds_151_1_106-128.pdf